

4

**CHEMICAL
RESEARCH,
DEVELOPMENT &
ENGINEERING
CENTER**

DTIC FILE COPY

CRDEC-CR-88083

**USING THEORETICAL DESCRIPTORS
IN STRUCTURE ACTIVITY RELATIONSHIPS
III. ELECTRONIC DESCRIPTORS**

**by Leland Y. Wilson
LOMA LINDA UNIVERSITY
Riverside, CA 92515**

**George R. Farni
RESEARCH DIRECTORATE**

August 1988

**DTIC
ELECTE
OCT 1 7 1988
S D
& H**



**U.S. ARMY
ARMAMENT
MUNITIONS
CHEMICAL COMMAND**

AD-A200 482

Aberdeen Proving Ground, Maryland 21010-5423

DISTRIBUTION STATEMENT A

**Approved for public release;
Distribution Unlimited**

22 4024 304

Disclaimer

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorizing documents.

Distribution Statement

Approved for public release; distribution is unlimited.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

ADA200482

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED			1b. RESTRICTIVE MARKINGS		
2a. SECURITY CLASSIFICATION AUTHORITY			3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.		
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE					
4. PERFORMING ORGANIZATION REPORT NUMBER(S) CRDEC-CR-88083			5. MONITORING ORGANIZATION REPORT NUMBER(S)		
6a. NAME OF PERFORMING ORGANIZATION Loma Linda University		6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION		
6c. ADDRESS (City, State, and ZIP Code) Riverside, CA 92515			7b. ADDRESS (City, State, and ZIP Code)		
8a. NAME OF FUNDING/SPONSORING ORGANIZATION CRDEC		8b. OFFICE SYMBOL (If applicable) SMCCR-RSP-C	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER DAAL03-86-D-0001		
8c. ADDRESS (City, State, and ZIP Code) Aberdeen Proving Ground, MD 21010-5423			10. SOURCE OF FUNDING NUMBERS		
			PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.
11. TITLE (Include Security Classification) Using Theoretical Descriptors in Structure Activity Relationships III. Electronic Descriptors					
12. PERSONAL AUTHOR(S) Wilson, Leland Y., and Famini, George R.					
13a. TYPE OF REPORT Contractor		13b. TIME COVERED FROM 87 Jul TO 87 Aug		14. DATE OF REPORT (Year, Month, Day) 1988 August	
15. PAGE COUNT 32					
16. SUPPLEMENTARY NOTATION COR: George R. Famini, SMCCR-RSP-C, (301) 671-2670					
17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number) Quantitative structure activity relationships Linear solvation energy relationships Hydrogen bonding, (continued on reverse)		
FIELD	GROUP	SUB-GROUP			
07	04				
19. ABSTRACT (Continue on reverse if necessary and identify by block number) Quantitative Structure Activity Relationships (QSAR) have been used successfully in the past to develop predictive equations for several biological and physical properties. A subset of QSAR, Linear Solvation Energy Relationships (LSER) have been used by Kamlet and Taft to predict and correlate a wide variety of solute/solvent interactions for over 100 properties. A major difficulty with LSER is the use of empirically derived descriptors. Work done in this laboratory has shown that it is possible to use theoretically determined descriptors such as molecular volume and a molecular polarizability in place of empirical descriptors. This paper describes the search for electronic descriptors to replace or to describe empirical terms related to hydrogen bonding.					
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT <input checked="" type="checkbox"/> UNCLASSIFIED/UNLIMITED <input type="checkbox"/> SAME AS RPT <input type="checkbox"/> DTIC USERS			21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED		
22a. NAME OF RESPONSIBLE INDIVIDUAL SANDRA J. JOHNSON			22b. TELEPHONE (Include Area Code) (301) 671-2914		22c. OFFICE SYMBOL SMCCR-SPS-T

UNCLASSIFIED

18. SUBJECT TERMS (continued)

Electronic descriptors

PREFACE

The work described in this report was authorized under Contract No. DAAL03-86-D-0001. This work was started in July 1987 and completed in August 1987.

The use of trade names or manufacturers' names in this report does not constitute an official endorsement of any commercial products. This report may not be cited for purposes of advertisement.

Reproduction of this document in whole or in part is prohibited except with permission of the Commander, U.S. Army Chemical Research, Development and Engineering Center, ATTN: SMCCR-SPS-T, Aberdeen Proving Ground, Maryland 21010-5423. However, the Defense Technical Information Center and the National Technical Information Service are authorized to reproduce the document for U.S. Government purposes.

This report has been approved for release to the public.

Accession For	
NTIS GRA&I	<input checked="checked" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

Blank

CONTENTS

	Page
1. INTRODUCTION	7
1.1 Quantitative Structure-Activity Relationships (QSAR) ...	7
1.2 Linear Relationships	8
1.3 Linear Solvation Energy Relationships	10
1.4 Example Of The Application Of LSER	11
1.5 Computational Chemistry And QSAR/LSER	12
1.6 Scope Of This Report	13
2. EXPERIMENTAL	14
2.1 Chemical Computational Facilities	14
2.2 Descriptor Models	15
2.2.1 Topological Electronic Index	16
2.2.2 Solvent Contact Area	17
2.2.3 Bond Energies	18
3. RESULTS	19
4. DISCUSSION	22
5. RECOMMENDATIONS	24
5.1 Quantum Mechanical	24
5.2 Statistical Mechanical	25
5.3 Data Analysis And Mathematical Models	25
LITERATURE CITED	31

Blank

USING THEORETICAL DESCRIPTORS IN STRUCTURE ACTIVITY RELATIONSHIPS

III. ELECTRONIC DESCRIPTORS

1 INTRODUCTION

Quantitative Structure-Activity Relationships (QSAR) have been used successfully in the past to correlate a variety of activities with many empirically derived and structure based descriptors (1 - 8). QSAR is a generalization of Linear Free Energy Relationships (LFER) and is based on work by Hammett in which he derived electronic descriptors for the dissociation of substituted benzoic acids and their esters (9). The basic tenet behind QSAR is that there is a connection between the microscopic (molecular structure) and macroscopic (empirical) properties such that it may be possible to predict empirical properties from the molecular structure. Molecular structure based properties, referred to as descriptors, can be calculated with computational chemistry techniques.

1.1 Quantitative Structure-Activity Relationships (QSAR)

The Quantitative Structure-Activity Relationship (QSAR) concept suggests that there can be a mathematical relationship between the molecular structure of a compound and its activity in a system. Several different structural descriptors have been used in QSAR equations. These range from experimentally determined π and σ to quantum mechanical energy levels. Activity, as used in QSAR, is defined as some chemical, physical or biological property. One example is the reciprocal of the dose of some substance required to produce a biological response. A chemical example is the distribution of a solute between two solvents.

A very important consequence of the QSAR idea is that if a mathematical structure-activity relationship can be found for a series of compounds the activity of some related compound can be predicted. Of the many possible mathematical relations a linear function is the simplest mathematically and conceptually and can be shown to be a valid possibility for QSAR.

1.2 Linear Relationships

The possibility of using linear relationships can be suggested by heuristic arguments which use kinetics and thermodynamics.

The familiar linear relationship between $\log K$, where K is the equilibrium constant, and the standard Gibbs free energy change, ΔG , for a given process is stated in (eqn.1).

$$\Delta G = - 2.30 R T \log K \quad (\text{eqn.1})$$

If there is a linear relationship between $\log A$, the logarithm of the activity, A , and $\log K$ then there is a linear relationship between $\log A$ and the Gibbs free energy change. A linear relationship between $\log A$ and $\log K$ can be obtained using an argument based on a reaction mechanism and an equilibrium constant expression (10).

A short explanation is based on the idea that the dose is related to the concentration of a reactant while the response of the system can be related to the concentration of a product. These concentrations can be related through an equilibrium concentration expression. The activity can

be taken as a function of a concentration. The equilibrium constant expression involves mathematical products of concentrations with some factors having negative exponents. Taking the logarithm of this mathematical product produces an algebraic sum of logarithms, a sum of logarithms is linear in those logarithms. As a result there can be linear relation between the $\log A$ and $\log K$.

Since there is a linear relationship between the $\log A$ and $\log K$ then (eqn.1) indicates that there is a linear relationship between $\log A$ and the gibbs free energy change. The relationship between ΔG , ΔH , and $T\Delta S$ as shown in (eqn.2) can, by simple substitution show that ΔH is linearly related to $\log A$.

$$\Delta H - T \Delta S = \Delta G \quad (\text{eqn.2})$$

This is the basis for the Linear Free Enthalpy Relationships (LFER). (Since the free energy and the enthalpy differ by $T \Delta S$ they have been interchanged freely.) LFER represents a subset of the QSAR concept.

A connection between the LFER and structure was made by Hammett through an analysis of dissociation constants for a series of benzoic acid derivatives (9). The difference between the Gibbs free energy changes for the dissociation of each acid was assumed to be due to the difference between their structures. This difference is associated with a functional group. Sigma is an empirical descriptor that is defined by the relation, $\Delta G(2) - \Delta G(1) = -2.30 R T \sigma'$ and its equivalent form, $\log K(2) - \log K(1) = \sigma'$. Hammett realized that this descriptor was, in fact, related to the electron withdrawing power of a particular attached group. This relationship and many generalizations made from it have formed the basis

behind LFER and subsequently, QSAR.

1.3 Linear Solvation Energy Relationships

Kamlet and Taft extended the LFER approach to the interaction of solutes and solvents (1,2). The interaction between a solvent and solute is a solvation process and the associated energy change is called a solvation energy. This type of Linear Free Enthalpy Relationship is called a Linear Solvation Energy Relationship (LSER).

Taft, Kamlet and co-workers correlated over 100 properties with what they call solvatochromatic parameters(1,2,3). The equation has the general form shown in (eqn.3) (4):

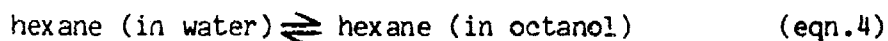
$$\text{LSER} = \text{cavity term} + \text{polarizability term} + \text{hydrogen bonding term} \\ + \text{intercept.} \quad (\text{eqn.3.})$$

The cavity term involves a volume (molar volume for solutes; Hildebrand solubility parameter for solvents). The polarizability and hydrogen bonding terms (one each for acidity and basicity) are expressed in terms of the solvatochromatic parameters which are obtained from UV-VIS spectroscopy.

The solvatochromatic parameters are empirical in nature; this means that the compound or a series of compounds have to be synthesized and the parameters measured. Once the parameters are measured the activity (dependent property, whatever the LSER represents) can be correlated with these parameters.

1.4 Example Of The Application Of LSER

An example of solute-solvent interaction is the distribution (partitioning) of hexane between octanol and water as represented by the following chemical equation:



The equilibrium constant expression is given by

$$K_{ow} = [\text{hexane(in water)}]/[\text{hexane(in octanol)}]$$

assuming one for the thermodynamic activity coefficients. This system provides a convenient model for LSER studies since the partition coefficient, K_{ow} , is easily measured and can be taken as the measure of activity for the system. Kamlet and Taft have log K_{ow} data for over 70 compounds and have done extensive analysis on this system (2). Furthermore the partition coefficient provides a parameter which models the lipophylic-hydrophylic blood-brain barrier.

One major difficulty with empirical descriptors is that the compound(s) in either the initial correlations, or subsequent predictions, have to be synthesized and the descriptors measured. This, to a large degree, detracts from the idea of property or activity prediction. The incorporation of descriptors directly derived from the structure of the molecule into the QSAR equations can potentially yield relationships where predictions of activities can be made without synthesis of the target compounds.

1.5 Computational Chemistry And QSAR/LSER

Replacing an equation using empirical descriptors with an equation using theoretical or computationally derived descriptors makes it possible to predict the activity (properties) of a compound *a priori*. In addition using theoretical or computational descriptors in place of empirical descriptors is a more direct application of the structure-activity relationship concept. Laboratory measurements are required in order to find the LSER empirical descriptors and require considerable lab space, time, and chemicals (some of which may be expensive or toxic). If molecular structure descriptors can be obtained by computational chemistry techniques then laboratory space, time and chemicals can be conserved. Furthermore the results of the computations can provide insights into the fundamental processes involved.

Calculated molecular descriptors have been used in structure-activity relations. For example Kaliszan and co-workers have used quantum chemical parameters in quantitative structure-retention relations (QSRR) involving gas chromatographic retention times (5,6). They were able to fit data with a linear relationship. Another example of QSRR is the work of Dunn and co-workers who have used molecular mechanics to obtain a molecular surface area which they correlated with the gas chromatographic retention index (7).

At the Chemometric/Biometric Modeling Branch of the Chemical Research, Development and Engineering Center (CRDEC) work has been done toward finding molecular parameters that can replace or correlate well with the terms in (eqn.3). The computational facilities there have been used to

demonstrate that the cavity term in (eqn.3) can be represented by the molecular volume (4). In addition the polarizability term in (eqn.3) can be well represented by a molecular polarizability parameter (7).

1.6 Scope Of This Report

The purpose of this paper was to search for a molecular parameter to represent the hydrogen bonding term in (eqn.3). Some of the descriptors that have been used will be described and suggestions will be made other descriptors.

The system and process employed is the distribution of hexane between two phases, octanol and water, as represented in (eqn.4) above. The partition coefficient, K_{ow} , will be taken as a measure of the activity of the hexane in this system.

Molecular descriptors which have been used in this laboratory to attempt to model hydrogen bonding basicities include the following: dipole moment; formal charges; a charge interaction which was a sum of products of pairs of charges divided by the square of the number of atoms; a charge-surface quantity which involved the double sum of atomic formal charges times the atomic areas divided by the double sum of the product of the atomic areas; and a surface weighted root-mean-square charge parameter which is a sum of the squares of the atomic formal charges times the atomic area divided by the total area. These quantities gave very poor statistical measures of fit as indicated by low values of the multiple correlation coefficients (11).

2 EXPERIMENTAL

The experimental work involved the use of the computational facilities at the Chemometric/Biometric Modeling Branch of the Chemical Research, Development, and Engineering Center (CRDEC).

A general procedure was instituted for developing predictive equations. First, molecular models are devised to describe hydrogen bonding. Then a set of compounds is selected. For each compound, the initial molecular geometry (bond distances, bond angles, dihedral angles and atomic connections) are set up and semi-empirical quantum chemical techniques are employed to optimize the geometry and produce electronic parameters. These geometric and electronic parameters are used to calculate other quantities required for the hydrogen bonding model. Finally a statistical (multiple linear regression) analysis, using log K_{ow} as the dependent variable, is done on the set of data that has been produced and the results are evaluated for statistical significance, inter-descriptor correlations and physical content.

2.1 Chemical Computational Facilities

Geometrical optimization was done using programs available in MOPAC, A General Molecular Orbital Package (12). Specifically employed were MNDO (minimum neglect of differential overlap) and AM1; these programs produce files containing the geometry, energy and electron population. AM1 is a program which is designed to give a good representation for hydrogen bonding (13).

Charge and size related parameters were calculated using programs from the Molecular Modeling, Analysis and Display System (MMADS) (14). This software package was developed at the Chemometric/Biometric Modeling Branch of the CRDEC. Two programs, CONOLLY and AREA, are available to calculate types of molecular surface areas. Another program, ELECTOP, was added to MMADS to calculate the topological electronic index. In addition STICK was used to display the input and optimized structures and to help set up the input geometries. ZINDO was used to produce an electron population from the optimized geometry. ZINDO uses the semi-empirical INDO/S (intermediate neglect of orbital overlap) methodology (14).

2.2 Descriptor Models

Several different approaches were used to model hydrogen bonding. A major assumption is that hydrogen bonding is related to size and electronic characteristics. Hydrogen bonding occurs with atoms that have a high electronegativity; this is correlated with small size and five or more valence electrons. Inherent in structure-property correlations is the idea that bond energies are determined by such size and charge properties. This suggests an alternate approach of directly calculating a measure of the bond energy by using energy parameters that result from the quantum chemical calculations.

The descriptors chosen are described in the following sections.

2.2.1 Topological Electronic Index

Kaliszan and co-workers defined a quantum chemical parameter called the topological electronic index, $T(E)$. It combines electronic and geometric descriptors. They used it in the correlation analysis of gas-liquid chromatographic retention indices (QSRR) (5,6). The topological electronic index is a measure of the differences in solute molecular constitution, shape and size and is defined by the following relation:

$$T(E) = \sum_{i \neq j} \text{abs}[fc(i)-fc(j)]/r(i,j)^2 \quad (\text{eqn.5})$$

The sum does not include terms with $i=j$ as this would give an indeterminate form. $T(E)$ is the topological electronic index; $fc(i)$ is the negative of the excess electron population density (formal charge). The $r(i,j)$ values are the internuclear distances.

A physical interpretation of the topological electronic index can be made by recognizing that each term involves a coulombic interaction and represents the magnitude of an electric field strength at a distance, $r(i,j)$, away from a charge of size, $(fc(i)-fc(j))$.

The distances and electron populations are obtained using quantum chemical calculations. The distances were obtained from the optimized geometry produced with the program, MNDO. The formal charges were calculated using the program, ZINDO, which is better for electronics (15). The program, ELECTOP, was written in FORTRAN (and incorporated into MMADS) for the purpose of calculating the topological electronic index.

Three other variations of the topological electronic index were also used. The first one, labeled $T(E)^2$, used the square of the difference in the formal charge in place of the first power in the definition, (eqn.5). The other two were the analogues of the first two with the formal charges replaced by the electron populations. These were represented by $T(E,P)$ and $T(E,P)^2$.

The topological electronic index was combined with the molecular volume and the solute-solvent contact area (SCA) to produce other descriptors. The solute-solvent contact area is described in the next section and the molecular volumes were described by Famini (4).

2.2.2 Solvent Contact Area

Since it is a geometrical descriptor a surface area, by itself, would not be an adequate electronic descriptor since it does not include the charge. In addition it correlates highly with the molecular volume. However, the part of the solute surface area which can be touched by a solvent molecule might provide a measure of the size contribution to hydrogen bonding. This solute-solvent contact area (SCA) is defined as the area of the surface of the solute molecule which a solvent molecule can touch (16). This would depend on the size (radius) of the solvent molecule and the shape of the solute molecule. The solute molecule is treated as sphere. The solvent contact area (SCA) was calculated with the program, CONOLLY (17), which has been incorporated within MMADS.

One choice of a function involving charge and size would be the product of the topological index and an area. It would have units of electrical charge only.

2.2.3 Bond Energies

Two approaches were used to model the hydrogen bond more directly. They are represented by the following processes:



and



Reaction enthalpies for the reaction involving H(+1) (eqn.6) were calculated using enthalpies of formation obtained with the program, MNDO, while those for the reaction involving H₂O (eqn.7) were calculated using the program, AM1. As in most quantum chemical calculations, the results apply only to gaseous state molecules where there are no inter-molecular interactions.

3 RESULTS

Each descriptor chosen was employed as the hydrogen bonding term in the following version of (eqn.3):

$$\log K_{ow} = a(\text{molecular volume}) + b(\text{polarizability}) + c(\text{descriptor}) + d \quad (\text{eqn.8})$$

The molecular volume and polarizability data were obtained from references (4) and (8). The set of data was analyzed using a multiple linear correlation analysis contained in the HASSLE statistical package which is available from the DEC Users Society (DECUS) library. The results for the goodness of fit are summarized by model below. Numerical values for selected descriptors are listed in table A.

The symbols employed are listed here.

$T(E)$ = topological electronic index

$T(E)^2$ = topological electronic index with squared charge

$T(E,P)$ = topological electronic index based on electron populations

$T(E,P)^2$ = topological electronic index with squared charge based on
electron populations

$McVol$ = molecular volume

SCA = solvent contact area

$\Delta H(6)$ = enthalpy of reaction with $H(+1)$ (eqn.6)

$\Delta H(7)$ = enthalpy of reaction with H_2O (eqn.7)

n = number of compounds, sample size

R = multiple correlation coefficient

SEE = standard error of the estimate

Descriptor	R	SEE
-----	-----	-----
n = 72		
T(E)	0.890	0.555
T(E)/McVol	0.894	0.545
SCA	0.667	0.908
T(E)/SCA	0.914	0.496
T(E)*SCA	0.858	0.625
n=38		
$\Delta H(6)$	0.351	1.25
T(E)^2	0.705	0.948
T(E)^2/McVol	0.736	0.905
T(E)/SCA	0.747	0.889
T(E,P)	0.514	1.15
T(E,P)/McVol	0.409	1.28
T(E,P)/SCA	0.475	1.18

The values for the enthalpy change in the reaction with H(+1) (eqn.6) were not analyzed statistically. That model was rejected for physical reasons. In some cases a hydrogen atom was extracted from the molecule to produce a hydrogen molecule and a positive (-onium) ion. That was a much stronger interaction than expected for hydrogen bonding.

Calculations with the modified version of the topological electronic index, T(E,P)², were not pursued after calculations with T(E,P) showed low multiple correlation coefficients.

Using the best hydrogen bonding descriptor, T(E)/SCA, (eqn.8) can be written explicitly as follows:

$$\log Kow = (2.97 \pm 0.25) * McVol/100 + (0.77 \pm 65.9) * PI * 10 + (-3.15 \pm 0.23) * T(E)/(SCA * 100) - 0.0739 \quad (eqn.9)$$

$$n = 72$$

$$R = 0.914$$

$$SEE = 0.496$$

PI is the molecular polarizability as described in reference (8).

4 DISCUSSION

The statistical quantities tabulated for the models in the results section form the basis for judging the quality of the descriptor models.

In the above list the best descriptor for hydrogen bonding is the topological electronic index divided by the solvent contact area. By themselves the topological electronic index and the solvent contact area give a low correlation coefficient. The topological electronic index divided by the solvent contact area does incorporate the charge-size concept associated with hydrogen bonding. The correlation index between the topological electronic index and the solvent contact area is 0.280; the correlation index between the ratio, topological electronic index/solvent contact area, and the solvent contact area alone is 0.028. This shows that the new parameter is less correlated with the denominator (0.028) but more correlated with the numerator (0.955). It appears, from a chemical sense, that in the ratio the charge effect is more important than the size effect.

The modified forms of the topological index do not seem to provide any improvement over that proposed by Kaliszan and co-workers. That suggests that the formal charge could be the best charge parameter to use.

The enthalpy change (related to the bond energies) gives the worst fit. That may indicate that the model of one solute molecule and one solvent molecule is not adequate to describe hydrogen bonding. It does not necessarily mean that a bond energy model should be discarded.

In order for the theoretical descriptor for hydrogen bonding to be considered acceptable or useable, a linear relationship must be generated that has approximately the same (or better) correlation as the empirical parameter, B, being replaced. With McVol, PI and T(E)/SCA as independent variables the multiple correlation coefficient for log Kow as the dependent variable is 0.914. With McVol, PI and B the multiple correlation coefficient is 0.957. Consequently the theoretical descriptor, T(E)/SCA, cannot replace the empirical hydrogen bonding descriptor, B, and achieve similar levels of precision. The descriptor, T(E)/SCA, does satisfy the condition that there should be little correlation between it and the other two variables in (eqn.8), molecular volume (0.255) and molecular polarizability (-0.196).

It is instructive to note some features of the topological electronic index despite its being an inadequate descriptor (by itself) for this investigation. Molecules with polar groups have higher values than those without. For the same type of functional group molecules with longer hydrocarbon chains tend to have smaller values. These observations are consistent with (eqn.5).

5 RECOMMENDATIONS

To continue the search for a molecular descriptor for hydrogen bonding the following models and their combinations might be promising. Some of these suggested models are small variations on the models used in this paper.

5.1 Quantum Mechanical

A) Divide the topological electronic index by the fraction of the total area that is accessible to the solvent. The fractional area might be a better size descriptor than the solvent contact area.

B) Divide the topological electronic index by the surface area of the atom with the most negative formal charge. The atomic surface area might be a better size descriptor than the solvent contact area.

C) Use the difference between the formal charge on the most electronegative atom when the molecule is near a water molecule and when it is isolated. This value would be a measure of the effect of having a hydrogen bond which would effect the electron population on atoms near the bond.

D) Use models consisting of the solute molecule and two or more water molecules. This is a more realistic model for a hydrogen bonding system.

E) Use the energy of the highest occupied molecular orbital (HOMO). This is some indication of the Lewis base strength particularly if the electrons can be considered as a classical lone pair.

F) Use other molecular orbital and their associated energies and geometry. The presence of a lone pair could indicate involvement in hydrogen bonding.

5.2 Statistical Mechanical

Employ statistical mechanical methods for models consisting of several water and solute molecules. The partition coefficient is a macroscopic property; therefore it represents the interaction of a large number of molecules. In principle such a system can be analyzed with quantum mechanics; however, the computations required are prohibitive because they involve interactions between the large number of atoms. Statistical mechanical calculations should be less lengthy because they involve interactions between the molecules which will be smaller in number than the atoms.

5.3 Data Analysis And Mathematical Models

The descriptors could further be analyzed using the statistical techniques associated with principle component analysis (18). A set of descriptors (and their combinations) can be analyzed to find those which seem to give the better correlations. It may be possible to find linear combinations which can be treated as new parameters. In this regard a statistical software package with spreadsheet capabilities is useful; sets of data can be readily manipulated in order to get various combinations.

Other mathematical models could be examined. While the linear model is convenient and can more easily be interpreted in terms of (eqn.3) it is possible that other functions may work well. For example a quadratic term for one or more of the descriptors might give an equation that better fits the data. The overall goal is to find molecular properties which can be quantitatively related to empirical properties. While doing this, it is also important to be able to attach physical significance to the descriptors, and mathematical terms and factors that occur in any equation that comes out of the analysis.

Table A
Values of Selected Molecular Descriptors

#	file	McVol	PI	II*	B	log Kow	T(E)	T/Vol *P3	SCA	T/SCA *P3	T*SCA	ΔH
units →	A ³						e/A ²	e/A ⁵	A ²	e/A ⁴	e	kcal
1	mjk01	119.0	.1000	-.08	0.00	3.90	0.535	4.50	68.5	7.81	35.6	-3.882
2	mjk02	106.7	.1045	0.00	0.00	3.44	0.126	1.18	58.2	2.16	7.33	-5.486
3	mjk03	99.6	.0997	0.00	0.00	3.11	1.230	12.35	57.0	21.6	70.1	-4.577
4	mjk04	100.4	.0997	-.08	0.00	3.39	0.510	5.08	61.1	8.34	31.2	-4.118
5	mjk05	89.2	.1025	0.00	0.00	3.00	0.184	2.05	54.8	3.36	10.1	-4.626
6	mjk06	82.4	.0986	-.08	0.0	2.89	0.453	5.50	55.0	8.24	24.9	-4.685
7	mjk07	101.3	.1204	.08	.1	2.60	0.923	9.11	84.4	10.9	77.9	-3.773
8	mjk08	91.6	.1159	.08	0.1	2.83	1.025	11.19	76.6	13.4	75.5	-5.77
9	mjk09	181.6	.1052	.14	.69	2.79	4.474	24.64	83.9	53.3	375.4	-4.854
10	mjk10	98.3	.1025	.39	.1	2.64	1.449	14.74	65.2	22.2	94.5	-7.578
11	mjk11	94.1	.1109	.29	.1	2.49	2.109	22.41	71.2	29.6	150.2	-3.845
12	mjk12	65.5	.0953	-.08	0.	2.3	0.435	6.54	47.8	9.10	20.8	-3.166
13	mjk13	86.2	.1162	.33	.1	2.29	0.955	11.08	75.5	12.7	72.1	-0.171
14	mjk14	80.9	.1010	.39	.1	2.04	1.256	15.53	59.8	21.0	75.1	-4.17
15	mjk15	131.5	.1020	.24	.71	1.45	4.266	32.44	66.7	64.0	284.5	-6.607
16	mjk16	117.1	.1021	.5	.65	1.38	3.293	28.12	68.7	47.9	226.2	-4.37
17	mjk17	79.6	.1196	.17	.70	1.30	3.066	38.52	63.2	48.6	193.8	-3.451
18	mjk18	107.3	.1022	.47	.46	1.20	5.744	53.53	67.6	84.9	385.3	-2.699
19	mjk19	100.1	.1001	.67	.50	0.91	3.176	31.73	61.5	51.6	195.3	-4.257
20	mjk20	90.5	.0995	.27	.47	0.89	3.534	39.05	60.5	58.4	213.8	-2.63
21	mjk21	80.4	.1002	.60	.38	0.88	2.371	29.49	55.7	42.5	132.1	-2.488
22	mjk22	105.6	.1064	.76	.53	0.81	2.919	27.64	60.0	48.6	175.1	-3.269
23	mjk23	88.9	.1018	.55	.45	0.73	5.484	61.69	60.9	90.0	334.0	-4.845
24	mjk24	96.2	.1010	.16	.7	0.70	2.881	29.95	55.7	51.7	160.5	-3.249
25	mjk25	64.3	.0979	.6	.38	0.59	2.100	32.65	49.8	42.2	104.6	-3.431
26	mjk26	78.6	.1025	.58	.55	0.46	2.889	36.76	52.7	54.8	152.3	-3.251
27	mjk27	130.4	.1038	.86	.78	0.34	6.691	51.31	67.3	99.5	450.3	-6.700
28	mjk28	81.0	.1009	.67	.48	0.29	3.562	43.97	56.1	63.5	199.8	-3.708
29	mjk29	181.4	.1127	.87	1.05	0.28	14.131	77.90	79.7	177.2	1126.2	-4.058
30	mjk30	70.8	.1005	.60	0.42	0.18	4.616	65.20	53.3	86.6	246.0	-5.605
31	mjk31	64.5	.1106	.83	.30	0.18	2.754	42.70	50.6	54.5	139.4	-2.681
32	mjk32	78.0	.1066	.14	.65	0.16	2.107	27.01	51.1	41.2	107.7	-18.286
33	mjk33	63.1	.0975	.75	.37	0.10	1.126	17.84	50.6	22.2	57.0	-2.828
34	mjk34	55.1	.0940	.27	.47	0.10	2.268	41.16	45.1	50.3	102.3	-1.561
35	mjk35	63.9	.0979	.71	.48	-.24	2.784	43.57	48.0	58.0	133.6	-2.543
36	mjk36	45.1	.0941	.75	.37	-.34	1.043	23.13	43.3	24.1	45.2	-1.447
37	mjk37	47.0	.1101	.85	.30	-.35	2.452	52.17	44.4	55.3	108.9	-3.224
38	mjk38	93.4	.1059	.88	.76	-.77	4.924	52.72	59.7	82.5	294.0	-4.477

Table A (continued)

39	mjk39	36.5	.0859	.40	.40	-.65	1.980	54.25	36.6	54.2	72.5
40	mjk40	54.2	.0927	.40	.45	-.30	2.693	49.69	45.1	59.7	121.5
41	mjk41	71.3	.0969	.40	.45	.28	2.782	39.02	50.1	55.5	139.4
42	mjk42	72.1	.0955	.40	.51	.05	3.422	47.46	50.1	68.3	171.4
43	mjk43	89.8	.0980	.40	.45	.99	3.015	33.57	57.7	52.2	174.0
44	mjk44	89.4	.0982	.40	.45	.76	3.127	34.98	56.6	55.2	177.0
45	mjk45	89.7	.0976	.40	.51	.61	3.562	39.71	54.4	65.5	193.8
46	mjk46	89.1	.0978	.40	.57	.36	4.108	46.11	52.9	77.7	189.4
47	mjk47	107.4	.0995	.40	.45	1.48	3.102	28.88	63.1	49.2	195.7
48	mjk48	106.8	.0996	.40	.51	1.21	3.674	34.40	61.1	60.2	224.5
49	mjk49	108.2	.0981	.40	.45	1.34	3.707	34.26	60.3	61.4	223.5
50	mjk50	106.5	.0996	.40	.57	.89	4.284	40.23	58.1	73.7	248.9
51	mjk51	107.2	.0992	.40	.51	1.28	3.926	36.62	59.6	65.9	234.0
52	mjk52	121.1	.1039	.40	.45	2.03	3.196	26.39	71.2	44.9	227.6
53	mjk53	125.9	.0992	.40	.51	1.48	4.530	35.98	62.9	72.0	284.9
54	mjk54	84.6	.1204	.59	.10	2.13	0.189	2.23	55.2	3.42	10.4
55	mjk55	143.1	.1204	.74	.41	2.64	5.595	39.10	79.5	70.4	444.8
56	mjk56	119.1	.1204	.90	.48	1.58	2.766	23.22	68.2	40.6	188.6
57	mjk57	131.4	.1252	.90	.33	2.28	3.176	24.17	71.1	44.7	225.8
58	mjk58	100.8	.1249	.92	.44	1.48	1.985	19.69	62.9	31.6	124.9
59	mjk59	101.8	.1209	.54	.11	2.69	0.623	6.12	61.7	10.1	38.4
60	mjk60	109.	.1237	.73	.22	2.11	3.280	30.09	66.5	49.2	218.4
61	mjk61	126.4	.1224	.69	.23	2.51	4.054	32.07	72.6	55.9	294.3
62	mjk62	144.3	.1208	.65	.23	3.18	4.150	28.76	80.0	51.9	332.0
63	mjk63	99.8	.1237	.71	.07	2.84	1.352	13.55	67.0	20.2	90.6
64	mjk64	105.8	.1280	.79	.06	2.99	1.085	10.26	71.2	15.2	77.3
65	mjk65	99.7	.1276	.90	.38	1.56	1.076	10.79	64.2	16.8	69.1
66	mjk66	101.7	.1307	.70	.14	1.85	2.684	26.39	70.0	38.4	187.9
67	mjk67	136.5	.1210	.41	.15	3.42	1.605	11.76	76.5	21.0	122.8
68	mjk68	118.9	.1212	.47	.13	3.20	1.102	9.27	69.7	15.8	76.8
69	mjk69	134.7	.1224	.90	.40	2.21	5.092	37.80	72.9	69.8	371.2
70	mjk70	149.1	.1257	.85	.35	2.61	3.236	21.70	77.3	41.9	250.1
71	mjk71	167.5	.1175	.85	.35	3.31	4.507	26.91	80.1	56.3	361.0
72	mjk72	137.	.1186	.88	.48	2.20	2.819	20.58	74.0	38.1	208.6

Notes: Table B contains the names of the compounds keyed to the numbers, #.

II* is the solvatochromatic polarizability descriptor.

B is the solvatochromatic hydrogen bonding basicity descriptor.

*P3 means times 10 raised to the positive third power.

e represents the atomic charge unit. A is the angstrom.

Table B

List of Compounds in Table A

1 hexane	39 methanol
2 cyclohexane	40 ethanol
3 2,2-dimethylpropane	41 n-propanol
4 pentane	42 2-propanol
5 cyclopentane	43 n-butanol
6 butane	44 2-methyl-1-propanol
7 tetrachloroethene	45 2-butanol
8 carbon tetrachloride	46 2-methyl-2-propanol
9 tripropyl amine	47 1-pentanol
10 1-chlorobutane	48 3-pentanol
11 1,1,1-trichloroethane	49 2,2-dimethyl-1-propanol
12 propane	50 2-methyl-2-butanol
13 trichloroethene	51 3-methyl-2-butanol
14 1-chloropropane	52 1-hexanol
15 triethyl amine	53 3,3-dimethyl-2-butanol
16 2-hexanone	54 benzene
17 N-methyl pyridine	55 ethyl benzoate
18 ethyl propanoate	56 acetophenone
19 2-pentanone	57 dimethyl aniline
20 diethyl ether	58 benzaldehyde
21 butanal	59 toluene
22 cyclohexanone	60 methoxy benzene
23 ethyl ethanoate	61 ethoxy benzene
24 ethyl dimethyl amine	62 propoxybenzene
25 propanal	63 chlorobenzene
26 tetrahydrofuran	64 bromobenzene
27 diethyl acetamide	65 cyanobenzene
28 butanone	66 nitrobenzene
29 hexamethyl phosphoramide	67 1,3,5-trimethylbenzene
30 methyl ethanoate	68 m-xylene
31 nitroethane	69 o-dimethoxy benzene
32 trimethyl amine	70 n,n-dimethyl amino toluene
33 propanenitrile	71 N,N-diethyl aniline
34 dimethyl ether	72 phenyl propanone
35 propanone	
36 ethanenitrile	
37 nitromethane	
38 dimethyl acetamide	

Blank

LITERATURE CITED

1. Kamlet, M.J., Abbound, J.L.M., Abraham, M.H., and Taft, R.W., J. Org. Chem. Vol. 48, p 2877 (1983).
2. Taft, R.W., Abraham, M.H., Famini, G.R., Doherty, R.M., Abbound, J.L.M., and Kamlet, M.J., J. Pharm. Sci. Vol. 74, p 807 (1985).
3. Kamlet, M.J., and Taft, R.W., Acta Chem. Scand. Vol. B39, p 611 (1985).
4. Famini, G.R., Using Theoretical Descriptors in Structural Activity Relationships I. Molecular Volume, CRDEC-TR-88031, U.S. Army Chemical Research, Development and Engineering Center, Aberdeen Proving Ground, MD, January 1988, UNCLASSIFIED Report.
5. Kaliszan, R., Chromatography, Vol. 2, p 19 (1987).
6. Osmialowski, K., Halkiewicz, J., and Kaliszan, R., J. Chromatogr. Vol. 361, p 63 (1986).
7. Dunn, W.J., III., Koehler, M., Stalling, D.L., and Schwartz, T.R., Anal. Chem. Vol. 58, p 1835 (1986).
8. Famini, G.R., Using Theoretical Descriptors in Structural Activity Relationships, II. Polarizability, CRDEC-TR-88031, U.S. Army Chemical Research, Development and Engineering Center, Aberdeen Proving Ground, MD January 1988, UNCLASSIFIED Report.
9. Hammett, L.P., J. Am. Chem. Soc. Vol. 59, p 96 (1937).
10. Franke, R., Theoretical Drug Design Methods, Elsevier, New York, NY, 1984.
11. Famini, G.R., Unpublished results.
12. MOPAC, A General Molecular Orbital Package, QCPE Program No. 455, available from Quantum Chemistry Program Exchange (QCPE), University of Indiana, Bloomington, IN 47405.
13. Dewar, M.J.S., Zoebisch, E.G., Healy, E.F., and Stewart, J.J.P., J. Am. Chem. Soc. Vol. 107, p 3902 (1985).
14. Leonard, J.M., A User's Guide to the Molecular Modeling, Analysis, and Display System (MMADS), CRDEC-TR-86039, U.S. Army Chemical Research, Development and Engineering Center, May 1986, UNCLASSIFIED Report.
15. Available from the Quantum Theory Project (QTP), University of Florida, Gainesville, FL.

16. Connolly, M.J., Science, Vol. 221, p 709.
17. Available from Quantum Chemistry Program Exchange (QCPE), University of Indiana, Bloomington, IN 47405.
18. Rohrbaugh, R.H, and Jurs, P.C., Anal. Chem. Vol. 57, p 2770 (1985).